OpenSM Release Notes

Rev 0.3.1

Mellanox Technologies

© Copyright 2004. Mellanox Technologies, Inc. All Rights Reserved.

OpenSM Release Notes

Document Number:

Mellanox Technologies, Inc.
2900 Stender Way
Santa Clara, CA 95054
U.S.A.
www.Mellanox.com

Tel: (408) 970-3400
Fax: (408) 970-3403

Mellanox Technologies Ltd
PO Box 586 Hermon Building
Yokneam 20692
Israel

Tel: +972-4-909-7200
Fax: +972-4-959-3245

Mellanox Technologies

1 Overview

This document describes the contents of OpenSM rev 0.3.1 release. OpenSM is an InfiniBand compliant Subnet Manager and Administrator, and runs on top of VAPI or OpenIB. It is provided in two flavors: a fixed flow executable named ***opensm***, and a configureable flow and policy Tcl extension named ***osmsh***. The two are accompanied by a testing application named ***osmtest***. Further documentation of the tools is provided in *OpenSM User's Manual, Document #2125SM*.

The document includes the following sections:

- "Overview" (page 3)
- "Known Issues" (page 6)
- "Unsupported IB Compliancy Statements" (page 7)
- "Bug Fixes" (page 9)
- "Main Verification Flows" (page 10)

1.1 New Features

No new features in this release.

1.2 Software Dependencies

OpenSM depends on the installation of either the Mellanox VAPI driver (pointed at by the MTHOME environment variable), or the OpenIB driver (pointed at by the TSHOME and MTHOME environment variables). The qualified driver versions are provided in Table 1, "Software Dependencies".

Table 1 - Software Dependencies

Task	Software	Supported Versions
HCA Driver & Special QP Management	Mellanox InfiniHost HCA Driver for Linux	3.1 and above
	OpenIB Stack	2.0.1

Mellanox Technologies

1.3 Supported Platforms

OpenSM has been qualified on the platforms and operating systems listed in Table 2 below.

Table 2 - Supported Platforms and Operating Systems

Architecture	OS	Kernel	GCC
X86-32	Red Hat Linux 7.3	2.4.18-10 3 (SMP)	2.96-110
	Red Hat Linux AS 2.1	2.4.9-e.12 (SMP and enterprise)	2.96-108.1
	Red Hat Linux 8.0	2.4.18-14 (SMP and bigmem)	3.2-7
	Red Hat Linux 8.0	kernel.org 2.4.22 (SMP)	3.2-7
	Red Hat Linux 8.0	kernel.org 2.4.23 (SMP)	3.2-7
	Red Hat Linux 9.0	2.4.20-8 (SMP and enterprise)	3.2-7
	Red Hat Enterprise Linux 3.0	2.4.21-4.EL big pages & kernel patch	3.2.3
	Red Hat Enterprise Linux 3.0	2.4.21-9.EL (SMP)	3.2.3
	SuSe SLES 8.0	2.4.19-64G-SMP	3.2
	SuSe 9.0	2.4.21-99-smp4G	3.3.1
X86-64 AMD	SuSE SLES-8 (AMD64)	2.4.19-NUMA	3.2.3
	Suse 9.0 Pro	2.4.21-102-smp	3.3.1
	Red Hat Enterprise Linux AS 3.0	2.4.21-4.EL	3.2.3
IA64	Red Hat Linux AS 2.1	2.4.18-e.40smp	2.96-112.7.2
	Red Hat Enterprise Linux AS 3.0	2.4.21-4.EL-SMP	3.2.3
X86-32 (PCI Express)	SLES 8.0	2.4.21-156-smp	3.2
	Red Hat 8.0	2.4.18-14smp - CA is MTLP25208 (Lion LP)	3.2
	Red Hat AS3.0	2.4.21-4.EL - CA is MTLP25208 (Lion LP)	3.2.3
	Red Hat 9.0	2.4.20-8smp	3.2

Mellanox Technologies

1.4 Supported FirmWare

The main task of OpenSM is to initialize InfiniBand devices. The devices and their corresponding firmware versions which were qualified using OpenSM are listed in Table 3 below.

Table 3 - Devices and Corresponding Firmware Qualified with OpenSM

Device	FW versions qualified
MT43132	5.2.0
MT47396	0.2.0
MT23108	3.1.0
MT25208	4.5

Mellanox Technologies

2 Known Issues

The major limitations of OpenSM are described in Table 4, “OpenSM Major Limitations”.

Table 4 - OpenSM Major Limitations

Unsupported Feature	Impacted Platforms	Impact
IB_MGT prevents more than one QP1 client on a port	All	OpenSM on top of VAPI will take control of QP1. ***osmtest*** and any other tool will need to be run through another port.
No Pkey update policy	All	OpenSM does not enable the configuration of Pkey Tables on the subnet.
IB "trusted" concept is unsupported	All	Queries that should be classified according to the trustworthiness of their sources will not be handled correctly.
No Service / Key associations	All	There is no way to manage Service access by Keys.
No SM to SM SMDB synchronization	All	Puts the burden of re-registering services, multicast groups, and inform-info on the client application.
Routing is not credit loop-free	All	Non-fat tree subnet topologies might deadlock in case of a heavy load.
No Direct Route Path FailOver	All	OpenSM uses direct route to communicate with the subnet devices. When a hardware failure prevents a DR-MAD to reach its destination, OpenSM fails to communicate with the entire subnet behind the failing node. If a FailOver mechanism were applied, another DR path could have been picked, thus providing a work-around the failing node.
Failure upon exit	X86-64 AMD only	OpensSM has a failure upon exiting under Mellanox’s InfiniHost HCA Driver.
OpenSM over OpenIB has not been checked on X86-32 (PCI Express)	X86-32 (PCI Express) only	OpenSM has not been checked yet over OpenIB stack on this platform. Will be checked in the next release of OpenSM.
Unsupported kernel	SuSE SLES-8 (ia64)	This kernel will be supported in the next release of OpenSM.
Unsupported kernel over Mellanox’s InfiniHost HCA Driver (OpenIBStack is supported)	Red Hat Enterprise Linux AS 3.0 (ia64)	This kernel will be supported in the next release of OpenSM.

Mellanox Technologies

3 Unsupported IB Compliancy Statements

The following table lists all the IB compliancy statements which OpenSM does not support. Please refer to IB specification for detailed information on compliancy.

Table 5 - OpenSM unsupported Compliancy Statements

Flow	Compliancy	Description
Authentication	C14-22	M_Key M_KeyProtectBits and M_KeyLeasePeriod shall be set in one SubnSet method. As a work-around, an OpenSM option is provided for defining the protect bits.
Authentication	C14-67	On SubnGet(SMInfo) and SubnSet(SMInfo) - if M_Key is not zero then the SM shall generate a SubnGetResp if the M_Key is matching or silently drop the packet if M_Key is not matching
Authentication	C15-0.1.23.1	PortInfoRecords shall always be provided with the M_Key component set to 0, except in the case of a trusted request, in which case the actual M_Key component contents shall be provided.
Authentication	C15-0.1.23.2	P_KeyTableRecords and ServiceAssociationRecords shall only be provided in responses to trusted requests.
Authentication	C15-0.1.23.4	InformInfoRecords shall always be provided with the QPN set to 0, except for the case of a trusted request, in which case the actual subscriber QPN shall be returned.
Event FWD	o13-17.1.2	If no permission to forward, the subscription should be removed and no further forwarding should occur
Handover	C14-37.1.2	Priority should be kept in non-volatile memory.
Handover	C14-38.1.1	Support AttributeModifier values in SubnSet(SMInfo). If the state transition requested is invalid - return with status code 7
Initialization	C14-24.1.1.5	GUIDInfo - SM should enable assigning Port GUIDInfo
Initialization	C14-44	If the SM discovers that it is missing an M_Key to update CA/RT/SW, it should notify the higher level.
Initialization	C14-62.1.1.11	PortInfo:VLHighLimit should match the configured VLArb on this port
Initialization	C14-62.1.1.12	PortInfo:M_Key - Set the M_Key to a node based random value
Initialization	C14-62.1.1.13	PortInfo:P_KeyProtectBits - set according to an optional policy
Initialization	C14-62.1.1.14	PortInfo:M_KeyLeasePeriod - set according to optional policy
Initialization	C14-62.1.1.22	GUIDInfo - SM should enable assigning Port GUIDInfo
Initialization	C14-62.1.1.24	SwitchInfo:DefaultPort - Not relevant, works only for random FDB
Initialization	C14-62.1.1.32	RandomForwardingTable
Multicast	o15-0.1.12	If the JoinState is SendOnlyNonMember = 1 (only), then the endport should join as sender only
Multicast	o15-0.1.13	If a Join request using unrealistic parameters is received, return ERR_REQ_INVALID

Mellanox Technologies

Table 5 - OpenSM unsupported Compliancy Statements

Flow	Compliancy	Description
Multicast	o15-0.1.8	If a request for creating an MCG with fields that cannot be met, return ERR_REQ_INVALID (SL FlowLabelTclass)
SA Query	C15-0.1.11	Query response should use only base lids (validation)
SA Query	C15-0.1.19	Respond to SubnGetMulti(MultiPathRec)
SA Query	C15-0.1.8.6	Respond to SubnAdmGetTraceTable
SA Query	C15-0.1.8.7	SubnAdmGetMulti SubnAdmGetMultiResp - Only in case of a MultiPath
SA Query		SubAdmGet/GetTable(GUIDInfo)
Services	C15-0.1.13	Reject ServiceRecord create, modify or delete if the given ServiceP_Key does not match the one included in the ServiceGID port and the port that sent the request
Services	C15-0.1.14	Provide means to associate service name and ServiceKeys

Mellanox Technologies

4 Bug Fixes

The following table lists all bugs fixed in this release of OpenSM.

Table 6 - Bug Fixes

	Description	Details
1	No traps were received over TopSpin stack.	Fix bug in the registration for traps over the TopSpin stack.
2	Health bit mechanism caused a crash of OpenSM and incorrect behavior..	Fix several bugs in the health bit handling mechanism.
3	When SM receives SwitchInfo with error during light sweep, it assumed the transaction ended ok.	If SwitchInfo is received with an error during light sweep - force a heavy sweep.
4	When connected in loop-back, SM sweeps only a heavy sweep.	Add support for light sweep when no switches are found.
5	In SA ServiceRecord queries - enabled setting a record with p_key zero, but failed on query of this record.	Fix bug in ServiceRecord query - if SR was initiated with p_key zero - do not perform p_key checks on requests for this service.
6	Bug in cl_event_wheel key creation	The key creation had a bug that caused a different creation on first call to function.

Mellanox Technologies

5 Main Verification Flows

OsmTest is the main automated verification tool used for OpenSM testing. Its verification flows are described in Table 7 below.

Table 7 - OsmTest Verification Flows

Test Flow	Verified Compliancy Statement
Inventory File	All port info, node info, and path records parameters
Service Record	- Register a service - Register another service (with a lease period) - Register another service (with service p_key set to zero) - Get all services by name - Delete the first service - Delete the third service.
Multicast Member Record	- Query of existing Groups (IPoIB) - BAD Join with insufficient comp mask (o15.0.1.3) - Create given MGID=0 (o15.0.1.4) - Create given MGID=0xFF12A01C,FE800000,00000000,12345678 (o15.0.1.4) - Create BAD MGID=0xFA. (o15.0.1.6) - Create BAD MGID=0xFF12A01B w/ link-local not set (o15.0.1.6) - New MGID with invalid join state (o15.0.1.9) - Retry of existing MGID - See JoinState update (o15.0.1.11) - BAD RATE when connecting to existing MGID (o15.0.1.13) - Partial JoinState delete request - removing FullMember (o15.0.1.14) - Full Delete of a group (o15.0.1.14) - Verify Delete by trying to Join deleted group (o15.0.1.14) - BAD Delete of IPoIB membership (no prev join) (o15.0.1.15)
Event Forwarding	- Register for information - Send a trap and wait for report - Unregister non-existing
Stress Testing	Flood the SA with queries from multiple channel adapters to check the robustness of the mechanism
Trap 64/65 Flow	Register to Trap 64-65, create traps (by disconnect/connect ports) and wait for report, then unregister.
Dynamic Changes	Dynamic Topology changes to test OpenSM adaptation & DB correctness

In addition to using OsmTest to verify the functionality of OpenSM, it is possible to further verify OpenSM by using it to setup an actual cluster. Once that is done, an automated check is performed where it is verified that the resulting port info and node info parameters are as expected. Furthermore, it is checked that the resulting routing tables are credit loop-free, and that they cover all point-to-point connectivity.

Another method for verifying OpenSM involves the use of an interactive shell which supports SM and SA configuration and querying.

Mellanox Technologies